

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

RhinAsthma patient perspective: A Rasch validation study

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1638430> since 2017-05-25T13:45:48Z

Published version:

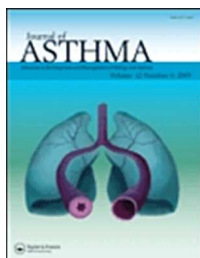
DOI:10.1080/02770903.2017.1316391

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)



RhinAsthma Patient Perspective: a Rasch validation study

Journal:	<i>Journal of Asthma</i>
Manuscript ID	LJAS-2017-0016
Manuscript Type:	Original
Keywords:	Quality of Life, Rhinitis/Sinusitis

SCHOLARONE™
Manuscripts

Abstract

Objective: In daily practice, Health Related Quality of Life (HRQoL) tools are useful for supplementing clinical data with the patient’s perspective. The availability of short and valid tools is crucial for improving usability by clinicians. A new HRQoL tool has recently been proposed for patients with asthma and allergic rhinitis: the RhinAsthma Patient Perspective – RAPP. The aim of the study was to evaluate the psychometric robustness of the RAPP using the modern IRT approach, to evaluate items scalability and test whether patients use the items response scale correctly.

Methods: One hundred and fifty-five patients (53.5% women, mean age 39.1 range 16-76) were recruited during a multicenter study in Italy. RAPP dimensionality and metric properties were investigated using IRT models. Differential item functioning (DIF) was evaluated for gender, age and asthma control test (ACT).

Results: The RAPP adequately fitted the Rating Scale model, demonstrating the equality of the rating scale structure for all items. All INFIT and OUTFIT statistics on items were satisfactory. The RAPP had adequate internal reliability and showed good ability to discriminate among different groups of participants. DIF analysis indicated that there were no differential item functioning issues for gender. One item showed a DIF by age and four items by ACT.

Conclusions: The psychometric evaluation performed using IRT models demonstrated that the RAPP met all the criteria for reliable and valid measurement. From a clinical perspective, this will allow physicians to confidently interpret scores as good indicators of QoL of patients with asthma.

Keywords: Validation, Rasch, DIF, Quality of life questionnaire, Assessment, Asthma, Rhinitis

For Peer Review Only

Introduction

The need for routine collection of Health Related Quality of Life (HRQoL) data and other Patient-Reported Outcomes (PROs) has recently been underlined [1-4]. PROs can be used in daily practice to supplement clinical data with the patient’s perspective, facilitate disease monitoring, improve patient engagement and sharing of decision-making [5-6].

A new HRQoL tool, the RhinAsthma Patient Perspective - RAPP [7], has recently been developed and validated for use in clinical settings for patients with asthma and allergic rhinitis. The RAPP has demonstrated adequate psychometric performance and clinical relevance. Moreover, it has the pragmatic characteristics that enhance uptake in routine care [8]: the questionnaire is self-administered, easy to complete, short, and it is simple to calculate and interpret the scores.

The psychometric properties of the RAPP questionnaire have only been tested using the classical test theory (CTT). Also when estimating a sophisticated model such as confirmatory factor analysis, the main aim of the CTT approach is to determine the capability of the items to measure a latent construct, and coherently it focuses on correlations between items and their properties.

In addition to the classical validity perspective, the modern IRT approach presents several desirable features [9-13]. First, IRT models directly analyze the relationship between the item level score (item difficulty) and the patient’s level of QoL (person ability), producing a joint measure of individuals and items. Second, IRT models produce a common metric scale starting from dichotomous or ordinal responses to a questionnaire. Moreover, they are suitable for testing whether the items response scale has been correctly interpreted and used by patients, i.e. whether respondents are able to discriminate between each point of the response scale of any item of the instrument.

In order to improve the reliability and validity of a short tool that can be used in daily clinical practice, the objective of the present study was to evaluate the psychometric adequacy of the RAPP using Rasch analysis, i.e. the current psychometric approach for conducting an in-depth analysis of the scaling properties of an assessment instrument.

Methods

Study design

Data were collected during a multicenter study that involved nine allergology and pulmonology centers in Italy. 155 participants, aged over 18 years and with a good knowledge of both the spoken and written Italian language, were recruited because of their clinical diagnosis of concomitant asthma and rhinitis, according to GINA and ARIA classifications.

The study was approved by the Ethics Committee of the Azienda Ospedaliera Universitaria San Martino, Genoa (Protocol number 602/2010 May 31, 2010, Register number 48/10) and was performed in accordance with the provisions of the 1995 Declaration of Helsinki (as revised in Edinburgh in 2000). All participants signed consent forms and received clear information about the research aims, the voluntary nature of participation, the anonymity of the data and results.

The patients completed the asthma control test (ACT) [14] and the RAPP questionnaire [7].

The RAPP scale included 8 items with a five point Likert-type response scale: never (0), rarely (1), sometimes (2), very often (4), always (5).

The RAPP data were analyzed to achieve the full potential of the IRT approach, that includes different kinds of models derived from the seminal work [15] of George Rasch (1960). The Rasch model assumes unidimensionality – all items in a scale measure the same single latent trait – and local independence – there should be no correlation between a person's response to

one item and those to any other items. His original model focused on dichotomous responses and stated that the probability that a person will affirm an item is a logistic function of the difference between the person's ability (i.e. the person's level of the measured trait) and the difficulty of the item (the level of the measured trait expressed by the item). Graphically this function is an S-shaped curve, ranging from 0-1 and called the item characteristic curve (ICC): the greater the ability (x-axis), the higher the probability (y-axis) of endorsing an item, passing through a flex point where the level of ability is equal to the difficulty of the item, so that the probability of affirming the item is equal to 0.5. The Rasch model provides a measurement of person ability relative to item difficulty, both of which are expressed in the same unit size (logits).

This model was extended [16], firstly to analyze polytomous items, i.e. with a graded response scale as in a Likert-type item. The partial credit model (PCM) [17] postulates that each item has a unique rating scale structure, while in the rating scale model (RSM) [18] all items share the same rating scale structure.

Statistical analyses

Statistical analyses were performed using IBM SPSS software v. 20.0 and WINSTEPS 3.72.3 (Beaverton, Oregon).

To assess the construct validity, item scalability, and the behavior of the answering scale of the RAPP questionnaire, both the PCM and RSM measurements were estimated using a joint maximum likelihood method. The likelihood-ratio (LR) test between the PCM and RSM specifications was used to choose between the two models, starting with the expectation to preserve the more parsimonious, that is the RSM. If the test was not significant then the RSM model was used [19-21].

The assumption of unidimensionality was tested by performing post-hoc principal component analysis of residuals, and a dimensionality score of ≤ 2 was taken as the rule of thumb (i.e., residuals did not contain other significant dimensions).

In order to evaluate how well each item fits the model, the information-weighted (INFIT) and outlier-sensitive (OUTFIT) indices were considered. These statistics measure information about responses given by people with an 'ability' level close to (INFIT) or distant from (INFIT) the item difficulty level: values for both of these in the 0.5-1.5 range are considered satisfactory [22].

Item discrimination was determined by calculating a point-measure correlation (i.e., a measure of the correlation between single item scores and the Rasch measure), considering values of ≥ 0.70 as acceptable.

The person separation index (PSI) and the reliability index (RI) indicate whether differences across patient measurements depend on actual differences in respondents' "ability" rather than measurement errors. By convention, values of PSI > 1.5 together with RI ≥ 0.70 were considered adequate. A value of RI=0.8 means that the instrument statistically differentiates between at least 3 groups (R=0.9 indicates the possibility of discriminating between 4 or more groups). [23, 24]

Lastly, since a good instrument should behave in the same manner for all respondents, a DIF (differential item functioning) analysis was performed to test measurement invariance, applying the same measurement models to different subgroups of participants. A difference of at least 0.5 logits between groups was noticeable, and indicated an item bias [25]. In this study, DIF was tested for gender, age and ACT.

Results

There were no missing data and all 155 cases were analyzed using WINSTEPS software.

Overall, the items in the RAPP adequately fitted both model specifications for polytomous items (PCM and RSM), but the RSM was more parsimonious and desirable, because of its assumption of equality of the rating scale structure for all items.

Based on a not significant likelihood ratio test ($\chi^2=38.8$; $df=21$; $p<0.05$), the RSM model was used.

Post-hoc principal component analysis of residuals yielded a value of 2, confirming that data respected the unidimensional assumption.

Table 1 shows the scaled RAPP; items are presented in order of QoL level measured by the RAPP, starting from the most arduous, in terms of the difficulty expressed. As the location decreased, the patient's QoL level increased, therefore reporting an intense frequency of "sleeping problems" in the two weeks preceding the visit corresponded to the lowest QoL.

TABLE 1 ABOUT HERE

Table 1. Scaled RAPP: items, location and fit statistics (Rating Scale Model).

All of the INFIT and OUTFIT statistics were in the 0.7-1.3 satisfactory range. The PT-Measure correlation values were similar and high for all items and only one item (item 4) revealed a slightly lower correlation value.

The RAPP had adequate internal reliability and showed good ability to discriminate among different groups of participants: PSI was 2.89 and RI .89.

The item locations ranged from -0.69 to +0.43 logits implying that the RAPP items are grouped in the middle of the logit scale, with no items covering both of the two extremes of the continuum of the person's level of QoL. This indicates that the scale does not work well with individuals with a medium or high QoL score.

The category probability curves of items are reported in Figure 1: the five response options for all items are well ordered and distributed and each one has a point at which it becomes the most likely response.

FIGURE 1 ABOUT HERE

Figure 1. Category probability curve.

DIF analysis indicated that there were no differential item functioning issues for gender. Only one item showed a small level of DIF by age as follows: difficulty in concentrating (.83 logits relatively easier in the group of older patients than in the 28-38-year-old group and .82 logits relatively easier in the group of older patients than in the 39-49-year-old group).

Four items showed a small level of DIF by ACT as follows: stuffy or runny nose, sneezing, or itchy nose (rated 0.97 logits relatively easier in the totally controlled group than in the uncontrolled group and rated 0.56 logits relatively easier in the well controlled group than in the uncontrolled group); Itchy, watery, sore and red eyes (rated 0.59 logits relatively easier in the totally controlled group than in the well controlled group, rated 1.22 logits relatively easier in the totally controlled group than in the uncontrolled group and rated 0.63 logits relatively easier in the well controlled group than in the uncontrolled group); Wheezing, cough, chest tightness, shortness of breath (1.79 logits relatively more difficult in the totally controlled group than in the well controlled group and 2.46 logits relatively more difficult in the totally controlled group than in the uncontrolled group and rated 0.67 logits more difficult in the well controlled group than in the uncontrolled group); and sleeping problems (.97 logits relatively more difficult in the totally controlled group than in the well controlled group and 1.49 logits relatively more difficult in the totally controlled group than in the uncontrolled group and rated 0.51 logits more difficult in the well controlled group than in the uncontrolled group).

Discussion

The objective of the present study was to evaluate the psychometric adequacy of the RAPP based on Rasch analysis. These findings indicate that the RAPP was a powerful tool and support the assertions of the classical test perspective on the psychometric properties of the RAPP [7]. Overall, the criteria for reliable and valid measurement were met. From a clinical perspective, this allows physicians to confidently interpret scores as good indicators of QoL of patients with asthma. Performance of scale items and response categories was analyzed in depth by exploiting the features offered by the Rasch models [10-13].

The RAPP revealed a unidimensional construct, as indicated by the fit statistics and PCA of the residuals; patients reacted to a five-point Likert rating scale in the desirable manner.

Rasch analysis allows items and persons to be measured on the same logit scale and enables simultaneous comparison of item difficulty and person ability. This feature is of great importance and is not available in the CTT perspective.

A well targeted tool has evenly spaced, well-fitting items that cover the full range of person abilities.

However, in this study a large floor effect with a small ceiling effect was observed for the RAPP, with the majority of our patients actually reporting a very good quality of life. It was evident that there are no items targeting persons with higher QoL, so the instrument would probably be more appropriate for more compromised patients. The presence of ceiling and floor effects is characteristic of HRQoL instruments since they have been constructed for use in a wide range of populations [26]. Thus, the findings may be a reflection of the sample's characteristics, i.e., healthy patients.

Our study shows the great value of Rasch analysis, which adds sophistication and refinement to traditional psychometric methods, and provides detailed diagnostic item-level data. In

conclusion, we found that the RAPP performed well on most aspects of the assessment: all items fit the construct, it has good discriminative ability, unidimensionality, the response scale functions as expected and there is a small amount of DIF which is in our opinion not of a magnitude to be problematic. The only serious problem for the overall scale was the poor targeting of item difficulty to person ability.

A limitation of this study is that the results were obtained on an Italian sample only, owing to the current availability of such data. Considering that validation of an instrument is a lengthy, even endless process [27], international studies will be performed to further test the psychometric properties of the RAPP.

Conclusions

This is the first study to use the Rasch model to evaluate the psychometric properties of the RAPP. The results uphold those of previous traditional testing and confirm that no changes to the questionnaire are required. This study provides further findings that support the use of the RAPP as a valid tool for the routine assessment of HRQoL in patients diagnosed with rhinitis and asthma.

References

1. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:167.

2. Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res*. 2012; 21:1305–14.

3. Porter I, Gonçalves-Bradley D, Ricci-Cabello I, Gibbons C, Gangannagaripalli J, Fitzpatrick R, Black N, Greenhalgh J, Valderas JM. Framework and guidance for implementing patient-reported outcomes in clinical practice: evidence, challenges and opportunities. *J Comp Eff Res*. 2016;5:507-19.

4. Eton DT, Beebe TJ, Hagen PT, Halyard MY, Montori VM, Naessens JM, Sloan JA, Thompson CA, Wood DL. Harmonizing and consolidating the measurement of patient-reported information at health care institutions: a position statement of the Mayo Clinic. *Patient Relat Outcome Meas*. 2014; 5:7-15.

5. Valderas JM, Kotzeva A, Espallargues M et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual. Life Res*. 2008;17:179-93.

6. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Serv Res*. 2013;13:211.

7. Braido F, Baiardini I, Stagi E, Scichilone N, Rossi O, Lombardi C et al. RhinAsthma patient perspective: a short daily asthma and rhinitis QoL assessment. *Allergy*. 2012;67:1443-50.

8. Kroenke K, Monahan PO, Kean J. Pragmatic characteristics of patient-reported outcome measures are important for use in clinical practice. *J Clin Epidemiol*. 2015;68:1085-92.

- 1
2
3 9. Cella DF, Lloyd SR, Wright BD. Cross-cultural instrument equating: current research and
4 future directions. In: Spilker B, editor, *Quality of Life and Pharmacoeconomics in Clinical*
5
6 *Trials* 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1996.
- 7
8
9 10. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement for the*
10
11 *Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.; 2001.
- 12
13 11. Smith EV Jr. Evidence for the reliability of measures and validity of measure
14 interpretation: a Rasch measurement perspective. *J Appl Meas*. 2001;2:281-311.
- 15
16
17 12. Tennant, A., McKenna, S. P. and Hagell, P. Application of Rasch Analysis in the
18
19 *Development and Application of Quality of Life Instruments*. *Value in Health*. 2004;7:S22-
20
21 S26.
- 22
23 13. Wu, M., & Adams, R. *Applying the Rasch model to psycho-social measurement: A*
24
25 *practical approach*. Melbourne: Educational Measurement Solutions; 2007.
- 26
27
28 14. Nathan RA, Sorkness CA, Kosinski M, Schatz M, Li JT, Marcus P et al. Development of
29
30 the asthma control test: a survey for assessing asthma control. *J Allergy Clin Immunol*.
31
32 2004;113:59-65.
- 33
34 15. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago:
35
36 University of Chicago Press; 1960 (Reprinted 1980).
- 37
38 16. Van der Linden WJ, Hambleton RK, editors. *Handbook of Modern Item Response*
39
40 *Theory*. New York: Springer; 1997.
- 41
42 17. Wright BD, Masters GN. *Rating Scale Analysis*. Chicago: MESA Press; 1982.
- 43
44 18. Andrich D. *Rasch Models for Measurement*. Series: Quantitative Applications in the
45
46 *Social Sciences* No. 68. London: Sage Publications; 1988.
- 47
48 19. Anderson, EB. A goodness of fit test for the Rasch model. *Psychometrika*. 1973;38:123-40.
- 49
50 20. Bock RD, Aitkin M. Marginal maximum likelihood estimation of item parameters:
51
52 Application of an EM algorithm. *Psychometrika*. 1981;46:443-59.
- 53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

21. Baker FB. Item Response Theory. New York: Marcel Dekker; 1992.

22. Wright BD, Linacre JM. Reasonable mean-square fit values. Rasch Measurement Transactions. 1994;8:370.

23. Wright BD. Reliability and separation. Rasch Measurement Transactions. 1996;9(4):472.

24. Wright BD. Interpreting reliabilities. Rasch Measurement Transactions. 1998;11(4):602.

25. Holland PW, Wainer H, editors. Differential Item Functioning. Mahwah, NJ: Lawrence Erlbaum Associates; 1993.

26. Upton P, Maddocks A, Eiser C, Barbes PM, Williams J. Development of a measure of the health-related quality of life of children in public care. Child Care Hlth Dev. 2005;1(4):409-15.

27. Cronbach LJ. Construct validation after thirty years. In: Linn RL, editor, Intelligence: Measurement, theory, and public policy. Urbana and Chicago: University of Illinois Press; 1989.

Table 1. Scaled RAPP: items, location and fit statistics (Rating Scale Model).

Items	Location	INFIT	OUTFIT	PT-Measure
	Logits (SE)	MNSQ	MNSQ	Corr.
5 Problems in sleeping (i.e., night awakenings)	0.43 (0.10)	1.07	1.05	0.70
6 Having to avoid certain areas and environments	0.20 (0.10)	0.85	0.83	0.74
3 Difficulty in concentrating	0.19 (0.10)	0.98	0.95	0.72
8 Limitations in performing some activities (i.e., working, studying, sports)	0.13 (0.10)	0.75	0.71	0.77
4 Wheezing, cough, chest tightness, shortness of breath	0.04 (0.10)	1.36	1.40	0.59
2 Itchy eyes, watery eyes, sore eyes, and eyes redness	-0.07 (0.10)	0.87	0.85	0.74
7 Having to take drugs	-0.21 (0.10)	1.11	1.08	0.73
1 Stuffy or runny nose, sneezing, or itchy nose	-0.69 (0.10)	1.02	1.05	0.72

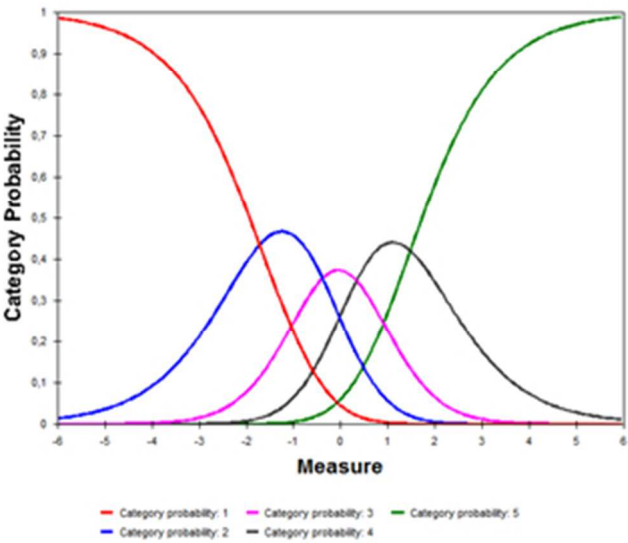


Figure 1. Category probability curve.

91x85mm (96 x 96 DPI)